Int J Life Cycle Assess (2014) 19:1854–1865
DOI 10.1007/s11367-014-0785-3

LCI METHODOLOGY AND DATABASES

Allocation in life cycle inventory: partial set of solutions to an ill-posed problem

Nathan B. Cruze · Prem K. Goel · Bhavik R. Bakshi

Received: 12 April 2013 / Accepted: 15 July 2014 / Published online: 13 August 2014

Abstract
Background, aim, and scope The primary aim of this paper is to indicate that partitioning allocation methods yields only a small subset of solutions to an ill-posed problem that has potentially infinitely many exact solutions. It will be shown that each of the existing partitioning methods arrives at just one particular solution from among infinitely many solutions of an *underdetermined system of linear equations*.
Materials and methods Some life cycle inventories fall into a class of functions called estimable functions in linear model framework, in which case they are invariant to allocation assumptions. This class of functions unites results described by Heijungs and Frischknecht (Int J Life Cycle Assess 3:321–332, 1998) and Heijungs and Suh (2002, Conjecture 1, p. 91). The inventories for non-estimable functions obtained through allocation are, in fact, derived under a set of additional implicit equality constraints called side conditions, often resulting in inventory results which differ greatly from one allocation to the next.
Results and discussions This paper explicates (1) identification of all estimable functions from any given technology matrix and (2) recovery of side conditions imposed on non-estimable functions through partitioning. These methods are illustrated in a simple example, and their relation to least squares techniques for allocation explored by Marvuglia et al. (Int J Life Cycle Assess 15:1020–1040, 2010) ;(Int J Agr Environ Inf Syst 3:51–71, 2012) are discussed.
Conclusions and outlook Recommendations are made that may lead to more meaningful ways to obtain additional data or include additional information in life cycle inventories in the future.

Keywords Allocation · Life cycle inventory · Matrix based LCI · Linear equations

Editorial responsibility: Nydia Suppen-Reynaga

N. B. Cruze · P. K. Goel
Department of Statistics, The Ohio State University, Columbus, OH 43210, USA

B. R. Bakshi (✉)
Lowrie Department of Chemical and Biomolecular Engineering, The Ohio State University, Columbus, OH 43210, USA
e-mail: bakshi.2@osu.edu

1 Introduction

An apparent source of subjectivity in the practice of life cycle assessment, allocation methods in the inventory phase have generated much debate. A considerable jargon has been developed to describe several methods for dealing with multifunctional processes, i.e., processes which yield more than one valuable output, treat more than one waste, or provide a combination of output and waste treatment functions ((Heijungs and Suh 2002), Definition 7, p.90). Two predominant methods for handling multifunctional processes are pursued in literature and in practice: partitioning and system boundary expansion ((Heijungs and Suh 2002), p.57).

Table 1 categorizes several references with regard to their area of study (theoretical discussion, case study, or policy concerns) and summarizes some of the important findings therein. This is by no means an exhaustive review on the subject of allocation. It is only intended to point out that (1) presently little theoretical understanding of allocation exists, (2) the effects of allocation are often assessed through application of multiple allocations on the same data in hopes of finding some agreement between methods, and (3) that debate persists in the literature as to which allocation principle is "right."

The primary aim of this paper is to indicate that partitioning explores only a small subset of solutions to a problem that is ill-posed in the sense that (1) it does not have just one solution, but potentially infinitely many solutions for some functional

Table 1 Summary table of reviewed allocation literature

Category	Reference	Contribution or Conclusion
Theoretical Developments	(Heijungs and Frischknecht 1998)	Noted several cases of exact solution via Moore-Penrose inverse
	((Heijungs and Suh 2002), Conjecture 1)	Asserted the equivalence of a solution via Moore-Penrose inverse to some partitioning
	(Marvuglia et al. 2010)	Investigated use of least squares techniques in lieu of partitioning. Recommended total least squares (TLS) method
	(Marvuglia et al. 2012)	Presented a generalized total least squares method for allocation
Case Studies	(Marvuglia et al. 2010)	Brick production: the five tabulated inventories often differed from each other by several orders of magnitude and sometimes by sign
	(Flysjö et al. 2011b); (Flysjö et al. 2011a)	Milk production (Sweden vs. New Zealand): life cycle GHG emissions attributed specifically to milk varied with choice of allocation. (85–98 % of total depending on partitioning criterion, 63–76 % by system expansion)
	(Cherubini et al. 2011)	Biorefinery: discussed a duality between system expansion and partitioning; noted that inventories of three co-products were most influenced by allocation criterion
	(Wang et al. 2011)	Transportation fuels: applied multiple allocations to various transportation fuels. Observed unusual reductions of well-to wheel GHG emissions by system expansion (as much as 130 % relative to diesel)
	Many others, e.g., (Cederberg and Stadig 2003), (Curran 2007), (Doluweera et al. 2011), (Silalertruksa and Gheewala 2011), (Wardenaar et al. 2012)	Each one observed considerable variations in respective inventories due to allocation assumptions
Allocation and Policy	(Weidema 2001); (Weidema and Schmidt 2010)	Advocate use of system boundary expansion versus partitioning
	(Cherubini and Strømman 2011)	Review noted that expert preferences were roughly evenly distributed among various allocation methods
	(Wardenaar et al. 2012)	Noted that different protocols recommended different means of allocation; debated policy implications

units and (2) these solutions are highly sensitive to the chosen allocation principle. Allocation methods simply choose a particular solution among infinitely many solutions of an underdetermined system of linear equations, possibly yielding a vast range of inventories. In fact, since one can easily inspect all possible solutions of this system, one need not settle for examining just a few solutions.

To this end, Section 2 presents the basics of linear model theory necessary to understand the allocation problem in generality using a simple illustrative example in Section 3. We reach the conclusions that every partitioning allocation procedure yields one of infinitely many exact solutions to an *underdetermined system* of linear equations and that a partitioning of multifunctional processes into single-function processes is equivalent to solving the system through an implicit choice of equality constraints, called side conditions, embodied by the use of some choice of a generalized inverse. In addition, some combinations of product inventories are *estimable linear parametric functions* in the sense that their estimates are invariant to the choice of allocation procedure. Strategies for identifying the estimable functions using projection matrices and other techniques are outlined. Furthermore, previous findings in the study of Heijungs and Frischknecht (1998) appertain to this study, and this insight resolves a conjecture about the construction of life cycle inventories via partitioning ((Heijungs and Suh 2002), Conjecture 1, p. 91). Finally, sometimes, desired inventories are not estimable functions, however, and the range of values these functions may take is determined by a simple formula. In Section 3, these ideas are further illustrated by continuing application to simple example presented in Section 2.1.3. We conclude in Section 4 that allocation is first and foremost an attempt to choose a meaningful solution to an *underdetermined* system of linear equations. In those instances when allocation is used, additional information *beyond the data* is required. It is incumbent on the practitioner to understand, disclose, and defend its validity.

2 Life cycle inventory and solutions of linear equations

By developing a deeper theoretical understanding of allocation, questions of why, how, and to what extent different methods of partitioning influence the final results of inventory may be better understood. Ultimately, it may lead to more meaningful ways of incorporating additional information into life cycle assessment (LCA) studies. In this section, we characterize the solutions of an allocation problem in terms of some basic results from solutions of linear equations and statistical linear model theory. A mathematical object called *the generalized inverse* of the unallocated technology matrix links inventories derived by partitioning multifunction processes with solutions of an underdetermined system of linear equations, with the individual process intensities as unknown variables.

A complete list of notation used throughout the remainder of the paper is provided in Section 1 of the supporting information. Where possible, we have retained conventions found in the study of Heijungs and Suh (2002).

2.1 Stating the ill-posed problem

2.1.1 Square, invertible technology matrix

Let p denote the number of products and n denote the number of processes in the system. The matrix formalism for life cycle inventory discussed in the study of Heijungs and Suh (2002) posits a linear relationship between production processes involved in the life cycle of a product (each process being represented as a column of the technology matrix, $\boldsymbol{A}_{p\times n}$) and some combination of net production, $\boldsymbol{f}$, in what might be considered the fundamental equation of life cycle inventory:

$$\boldsymbol{As} = \boldsymbol{f}. \tag{1}$$

Often, the basic goal of the inventory step is to determine the effects of just one unit of the j^{th} product, say $\boldsymbol{f}_j=[0\ 0\ \ldots\ 1\ \ldots\ 0]^T$. When the technology matrix $\boldsymbol{A}$ is square (i.e., $p=n$) and of full rank, the vector of scale factors, $\boldsymbol{s}_j$, at which each process must be operated to produce the target output, may be obtained by inverting Eq. 1:

$$\boldsymbol{s}_j = \boldsymbol{A}^{-1}\boldsymbol{f}_j. \tag{2}$$

It is worth noting that $\boldsymbol{s}_j$ is the j^{th} column of $\boldsymbol{A}^{-1}$.

Without loss of generality, one can consider just one environmental intervention at a time, denoted by a vector $\boldsymbol{b}$. Assuming that the impacts of production are linearly related to the scale at which each process is operated, it follows from Eq. 2 that the life cycle inventory, for one unit of the j^{th} product, g_j, is expressed

$$g_j = \boldsymbol{b}^T\boldsymbol{s}_j. \tag{3}$$

As explained in the supporting information, a vector of intensity factors, denoted $\boldsymbol{r}$, may be obtained as

$$\boldsymbol{r}^T = \boldsymbol{b}^T\boldsymbol{A}^{-1}. \tag{4}$$

This vector is essentially a list of per unit life cycle intensities for each product present in the system, and $\boldsymbol{r}^T$ is a row of the intensity matrix as defined in the study of Heijungs and Suh (2002). Post-multiplying both sides of Eq. 4 by $\boldsymbol{f}_j$ recovers Eq. 3; in words, g_j is the j^{th} component of the intensity vector $\boldsymbol{r}$.

Trivially, Eq. 4 can be expressed as

$$\boldsymbol{A}^T\boldsymbol{r} = \boldsymbol{b}. \tag{5}$$

Equation 5 suggests that the emissions observed at each process, indexed by *rows* of $\boldsymbol{A}^T$, are linearly related to exchanges of inputs and outputs in each process, and in particular, it asserts that the direct emissions at the j^{th} process are equal to the life cycle emissions of products made in that process, less the life cycle emissions of inputs into that process.

This leads to an equivalent calculation for life cycle inventories in which the intensity vector $\boldsymbol{r}$ is a solution of Eq. 5, and each component of the life cycle inventory of some arbitrary functional unit $\boldsymbol{f}$ may be expressed as $g=\boldsymbol{f}^T\boldsymbol{r}$. Given a full-rank technology matrix, this solution $\boldsymbol{r}$ is uniquely determined. However, when $p>n$, this is not true, as discussed in the next section.

2.1.2 Systems with more products than processes

Whenever the number of products exceeds the number of processes, i.e., $p>n$, the collection of processes *must* contain at least one multifunctional process. The unallocated technology matrix is rectangular. *Even in the presence of multifunctional processes, inventories derived from Eqs. 1 and 3 or equivalently taken directly from Eq. 5 still share an important relationship based on the singular value decomposition (SVD) of* $\boldsymbol{A}$. (See, e.g., ((Meyer 2001), p. 411–412).)

Given more products than processes, the fundamental equation shown in Eq. 1 is an *overdetermined* system of linear equations with scale factors as variables, having fewer variables than the number of equations. It is well known that such a system of equations may have an unique solution provided it is consistent, i.e., $\boldsymbol{f}$ is a linear combination of columns of the matrix $\boldsymbol{A}$; otherwise, the system is not solvable in that it has no solution at all. This lack of a solution to Eq. 1 has marked a point of departure for most of the allocation literature.

Typically, the LCA community has resorted to partitioning multifunctional processes into virtual, single-function processes, assigning shares of inputs and environmental interventions by some partitioning weights to create a square, invertible technology matrix $\boldsymbol{A}_*$ (of dimension $p\times p$) and corresponding vectors of interventions $\boldsymbol{b}_*$. The overdetermined

problem Eq. 1 is replaced by full-rank problem, and an inventory for the stated functional unit is extracted by

$$\boldsymbol{r}_*^T\boldsymbol{f} = \boldsymbol{b}_*^T\boldsymbol{A}_*^{-1}\boldsymbol{f}. \tag{6}$$

While resorting to partitioning, practitioners are actually solving the analogue of Eq. 5 for the allocated system:

$$\boldsymbol{A}_*^T\boldsymbol{r} = \boldsymbol{b}_* \tag{7}$$

and interpreting the linear combination of intensities $\boldsymbol{f}^T\boldsymbol{r}_*$ as the inventory for the stated functional unit. *In this expanded system, the solution* $\boldsymbol{r}_*$ *becomes unique*, but the choice of allocation will usually influence the result. The partitioning weights are essentially free variables in the creation of an allocated system as in Eq. 7, and the variety of partitioning criteria pursued in the literature (mass, energy, economic value, etc.) attests to the existence of many solutions.

Quite recently, Marvuglia et al. (2010, 2012) explored the use of a variety of least squares techniques, namely ordinary least squares (OLS), data least squares (DLS), and total least squares (TLS), in an attempt to find a possible approximate solution to Eq. 5 without resorting to allocation. Note that all these least squares problems can be written succinctly as variants of the TLS optimization problem (DeGroat and Dowling 1993):

$$\begin{aligned} &\min_{\boldsymbol{s},\Delta_A,\Delta_f} \left\| \left[\Delta_A \middle| \Delta_f\right] \right\|_F \\ &\text{subject to } (\boldsymbol{A}+\Delta_A)\boldsymbol{s} = \boldsymbol{f}+\Delta_f \end{aligned} \tag{8}$$

The idea of the above program is to make the smallest corrections to the technology matrix and functional unit (in the sense of the Frobenius norm) and report the scale factors which satisfies the constraints. Having obtained an approximate solution of Eq. 1, as a vector of scaling factors satisfying Eq. 8, the remaining inventory algebra is then carried out as in Eq. 3. (The reader is referred to (Van Huffel and Vandewalle 1991; Marvuglia et al. 2010; Marvuglia et al. 2012) and the supporting information for more on the computation of OLS (corresponding to Δ_A=0), DLS (corresponding to Δ_f=0), and TLS scale factors; an *R* script for the example in this paper has been provided with the electronic supporting information.)

We note that whenever $p>n$, the corresponding Eq. 5 for the *unallocated network data* is an *underdetermined* system of \(n\) linear equations i.e. one equation per process, with p product life cycle intensities as unknown variables , in contrast to the overdetermined system of Eq. 1, with \(n\) scale factors as unknown variables, having \(p\) equations, i.e. one equation per product. In such undetermined systems, *infinitely many solution vectors (i.e., intensity vectors) exist*. Clearly, not all individual product intensities in a solution vector $\boldsymbol{r}$ will be uniquely determined and this leads to the ill-posed nature of the problem stated in its simplest form: *How can one find a meaningful solution to Eq. 5 when the system involves a greater number of products than processes?* To illustrate the link between partitioning allocation and the underdetermined system of equations in Eq. 5, the following example is useful.

2.1.3 An illustrative example

A collection of processes in a waste-to-energy system as shown in Fig. 1 operates together exchanging four distinct products. The separator process engages in the separation of a combined waste stream into two components. In addition, a fuel oil production module and electricity production module produce energy for use within the system and outside the system boundary. Measurements of streams of two types of common pollutants are also taken at each of the processes.

The "recipes" shown in Fig. 1 for these products are embodied by the columns of the technology matrix shown in Eq. 9

$$A = \begin{array}{c} \begin{matrix} \text{Separation process} & \text{Fuel production} & \text{Electricity production} \end{matrix} \\ \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 1 & -1 \\ -2 & -3 & 6 \end{bmatrix} \end{array} \begin{array}{l} \text{kg waste 1} \\ \text{kg waste 2} \\ L \text{ fuel oil} \\ \text{kWh electricity} \end{array} . \tag{9}$$

Operating each process at scale $\boldsymbol{s}=(1\ 1\ 1)^T$, the system can produce a net output of 1 kWh electricity with a corresponding matrix of environmental interventions (with both emissions measured in grams) associated with each of the three reporting units is given by

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{b}_{CO_2}^T \\ \boldsymbol{b}_{NO_x}^T \end{bmatrix} = \begin{bmatrix} 8 & 9 & 21 \\ 15 & 14 & 3 \end{bmatrix}. \tag{10}$$

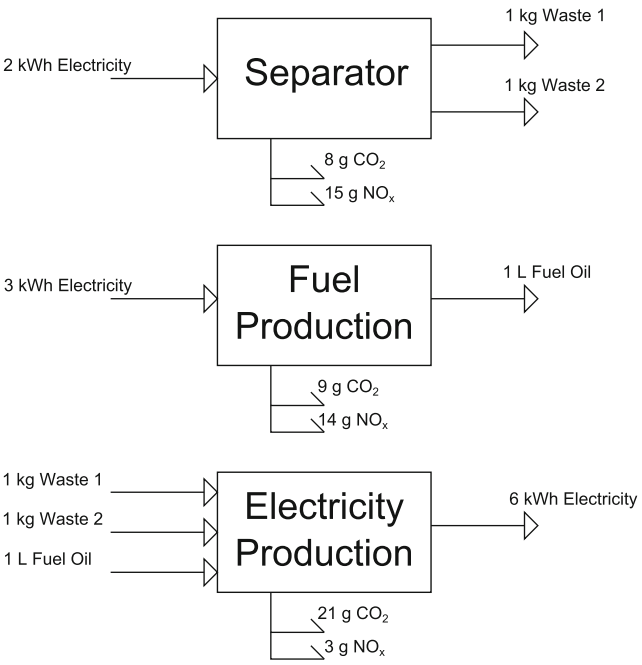

Fig. 1 Processes involved in the life cycles of four products

One may wish to evaluate the life cycle intensities of any of the products that may be present within the system boundary with respect to emissions of CO_2 and NO_x, formulating a partitioned technology matrix according to some weight, w, as shown in Eq. 11:

$$\boldsymbol{A}_* = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ -2w & -2(1-w) & -3 & 6 \end{bmatrix}. \tag{11}$$

The choice of w might be derived on the basis of any number of criteria, for example the relative shares of mass or economic value of streams leaving the separator. In the same manner, the allocation weight is applied to each vector of emissions, wastes, and resources. For the toy data, the allocated matrix of environmental interventions, $\boldsymbol{B}_*$, may be defined as follows:

$$\boldsymbol{B}_* = \begin{bmatrix} \boldsymbol{b}^T_{CO_2*} \\ \boldsymbol{b}^T_{NO_x*} \end{bmatrix} = \begin{bmatrix} 8w & 8(1-w) & 9 & 21 \\ 15w & 15(1-w) & 14 & 3 \end{bmatrix}. \tag{12}$$

Note that taking the sum of the first two columns in both Eqs. 11 and 12 retrieves the first column in each of the original, unallocated matrices.

Given Eqs. 11 and 12 above, symbolic computation of $\boldsymbol{B}_*\boldsymbol{A}_*^{-1}$ yields an intensity matrix as a function of an arbitrary allocation weight, w:

$$\begin{bmatrix} \boldsymbol{r}^T_{CO_2*} \\ \boldsymbol{r}^T_{NO_x*} \end{bmatrix} = \begin{bmatrix} 84w & 84(1-w) & 123 & 38 \\ 79w & 79(1-w) & 110 & 32 \end{bmatrix}. \tag{13}$$

The life cycle inventories for 1 L fuel oil and 1 kWh electricity (third and fourth columns of Eq. 4, respectively) are unchanged by the choice of w. Only the first two inventories are dependent on those weights, but note that *their sum is fixed*, whatever the choice of weights. From the available data, the sum of inventories of these two inventories is fixed at a total of 84 g of CO_2 and 79 g of NO_x. In this case, stating a choice of weight w to assign inputs, emissions, wastes, and resources to waste 1 would divide the sum of the inventories for both products accordingly.

Given Eqs. 11, 12, and 13 above, the reader can formulate the corresponding Eq. 7 for CO_2 and NO_x and inspect that they will hold for a choice of w. For any choice of w, we observe that each $\boldsymbol{r}_*$ must also satisfy the corresponding Eq. 5. In this example, $\boldsymbol{r}_{CO_2*}$ from Eq. 13 solves

$$\begin{bmatrix} 1 & 1 & 0 & -2 \\ 0 & 0 & 1 & -3 \\ -1 & -1 & -1 & 6 \end{bmatrix} \begin{bmatrix} 84\,w \\ 84(1-w) \\ 123 \\ 38 \end{bmatrix} = \begin{bmatrix} 8 \\ 9 \\ 21 \end{bmatrix} \tag{14}$$

for *all real* w; the typical partitioning allocation would constrain $0 \le w \le 1$. This example demonstrates that each vector of intensities obtained by partitioning allocation (by the specification of w) is also a solution of an underdetermined system of equations based on the *unallocated* data.

The link between Eq. 5 for unallocated data and Eq. 7 for *any* partitioning of that data is the key to fully understanding *why, how, and to what extent* inventory results may differ from one choice of allocation to the next. In the next section, we will show that this same $\boldsymbol{r}_*$ is only a particular solution of the underdetermined system of equations Eq. 5 for the unallocated data. Partitioning is simply a device used to obtain a solution from a system with fewer equations than unknowns. Known results from linear model theory, specifically the theory of constrained least squares, fully explain this link. The theory presented in the next section resolves Conjecture 1 proposed in the study of Heijungs and Suh (2002) and unifies it with the earlier results in the study of Heijungs and Frischknecht (1998). It also allows one to understand that the least squares techniques pursued in the study of Marvuglia et al. (2010, 2012) are in fact a type of partitioning allocation and all partitioning allocations can be described in terms of equality-constrained solutions of Eq. 5 of the unallocated data.

2.2 Partitioning and the less-than-full-rank linear model

We now consider the situation $p > n$. If rank($\boldsymbol{A}$)=n, that is, all n columns of $\boldsymbol{A}$ are linearly independent so that no process can be expressed as a linear combination of others, then Eq. 5 is a consistent, underdetermined set of linear equations, with infinitely many *exact* solutions.

Note that these exact solutions are automatically *least squares solutions* of Eq. 5 in the sense that all observations have zero residuals, so the error sum of squares has the minimum possible value of zero. Any least squares solution, say $\widehat{\boldsymbol{r}}$, to Eq. 5 must solve the set of p normal equations (see for example (Kutner et al. 2004)):

$$\left(\boldsymbol{A}\boldsymbol{A}^T\right)\boldsymbol{r} = \boldsymbol{A}\boldsymbol{b}. \tag{15}$$

It is known that the two underdetermined systems of equations shown in Eqs. 5 and 15 are equivalent to each other, i.e., all solutions of Eq. 5 are solutions of Eq. 15 and vice versa.

Definition 1: For a given matrix $\boldsymbol{D}$, any matrix $\boldsymbol{G}$ which satisfies the relation $\boldsymbol{D}\boldsymbol{G}\boldsymbol{D}=\boldsymbol{D}$ is said to be a generalized inverse (g-inverse) of $\boldsymbol{D}$ Note that if $\boldsymbol{D}$ has full rank, then its inverse is the unique generalized inverse. In the present context, rank($\boldsymbol{A}$)=n is smaller than the number of model parameters, so that the normal equations can be solved by a choice of non-unique *generalized inverse* of $\boldsymbol{A}\boldsymbol{A}^T$, denoted $(\boldsymbol{A}\boldsymbol{A}^T)^-$. One particular solution may be obtained as

$$\widehat{\boldsymbol{r}} = \left(\boldsymbol{A}\boldsymbol{A}^T\right)^+\boldsymbol{A}\boldsymbol{b} = \left(\boldsymbol{A}^T\right)^+\boldsymbol{b}, \tag{16}$$

where $(AA^T)^+$ and $(A^T)^+$ denote the Moore-Penrose pseudoinverses of (AA^T) and A^T, respectively. Note that in this less-than-full-rank linear model, $\hat{r}$ is but one possible solution of the normal equations. Choosing different generalized inverse, another solution could be

$$r_* = \left(AA^T\right)^- Ab = \left(A^T\right)^- b. \tag{17}$$

In the rank deficient case, the solution obtained by the pseudoinverse is known to have important geometric interpretations, yielding the "shortest length" solution among all possible solutions (i.e., $\sqrt{\hat{r}^T\hat{r}} \leq \sqrt{r_*^T r_*}$ among all other solutions r_* ((Golub and Kahan 1965), p.205)), a fact which makes it a meaningful solution in some applications (Kalaba et al. 1995). It is known that a choice of *any other g-inverse*, as in Eq. 17, may yield a substantially different solution of the normal equations. (See, e.g., (Ravishanker and Dey 2002), or any intermediate regression or econometrics text. (Harville (1997) and Ben-Israel and Grenville (2003) provide excellent references on the subject of g-inverses.)

Note that the set of all possible solutions R_b to the underdetermined system of linear equations Eq. 5 can be obtained from any particular solution plus arbitrary solutions to the homogeneous system $A^T r=0$, that is:

$$R_b = \left\{ r \middle| r = \left(AA^T\right)^- Ab + (I - P_A)h \right\}. \tag{18}$$

In Eq. 18 above, the term $(AA^T)^- Ab$ represents one particular solution (a vector of intensities) satisfying Eq. 5 and h represents an arbitrary vector of dimension p; P_A defined by

$$P_A = A\left(A^T A\right)^- A^T \tag{19}$$

is the unique orthogonal projection matrix onto the column space of A which finds the vector in the column space of A (row space of A^T) which is closest—in the sense Euclidean distance—to any arbitrary vector of length p. Note that this projection matrix does not depend on the choice of $(AA^T)^-$ used to solve Eq. 15. The term $(I - P_A)h$ represents an arbitrary solution to $A^T r=0$, the homogeneous system of equations. For non-zero columns of $(I - P_A)$, the corresponding components of r are changed as h ranges through all possible values. (See ((Meyer 2001), p. 64, p. 429) for more details.)

Equation 18 means that any solution of Eq. 5 can be expressed as the translation of some other solution (Lay 2003). We claim here that the set, C_b, of solutions derived via partitioning allocations is a subset of all possible solutions in R_b and offer more formal proof in Section 2.2 2. Argument about the quality of solutions found by partitioning amount to debating which, if any, solutions in this subset C_b are meaningful, but this still leaves out other solutions from consideration.

2.2.1 Identification of estimable functions of product inventories

Whether inventory calculation is considered as the direct estimation of coefficients of a regression model or as multiplication of emissions, wastes and resources by carefully chosen scaling factors, it is crucial to consider which inventories are uniquely *estimable* from the given data. Even though the individual product inventories contained within a solution vector of Eq. 5 may vary drastically with the choice of partition, some linear combinations of these inventories are preserved, regardless of the choice of g-inverse; these linear combinations are known as estimable functions in less-than-full-rank linear models (See, e.g., (Ravishanker and Dey 2002)).

Definition 2: Given the functional unit f, a linear combination of product life cycle inventories $f^T r$ is known to be estimable if and only if the row vector f^T can be written as a linear combination of rows of A^T, i.e., $f=As$ for some vector of scaling factors ((Ravishanker and Dey 2002), p. 114).

The very definition of estimability requires an exact solution to the fundamental equation! When Eq. 1 is consistent (i.e., has one unique solution), the inventory for the stated functional unit is an estimable linear function of model parameters, and the choice of g-inverses used to solve Eq. 5 or Eq. 15 becomes irrelevant (Rao 1962; Searle 1965; Ravishanker and Dey 2002).

Heijungs and Frischknecht ((1998), p. 329) intuited this with their case V scenario, in which scale factors for the inventory of a *co-product* could be computed exactly using the pseudoinverse. The authors reached an appropriate conclusion; that is, it would not be necessary to initiate an allocation procedure when the desired inventory could be computed exactly using the pseudoinverse of the technology matrix. Because the system of equations defining Eq. 1 in that example was *consistent*, using any other g-inverse of the technology matrix would, in fact, solve for the same unique solution. If there are no scale factors s satisfying Eq. 1, that is, the stated functional unit is not already a linear combination of columns of the unallocated technology matrix, then the inventory of desired functional unit is not an estimable function and the resulting inventory will inevitably be affected by the choice of g-inverse.

Note that very practical and low computational cost checks for estimability could be easily implemented for all life cycle inventories. Any functional unit of interest, f, can be *uniquely* written as a sum of its orthogonal projection, $P_A f$, and the discrepancy vector, $(I - P_A)f$. Thus,

$$f = P_A f + (I - P_A)f, \tag{20}$$

The projection $P_A f$ describes the portion of an intended functional unit f that can be explained by the columns of the

original data, $\boldsymbol{A}$, and this is the part which is estimable from the given data. Projection matrices offer an effective way to identify *all estimable functional units* for a given technology matrix. Grouping demands for a single unit of each product $[\boldsymbol{f}_1|\boldsymbol{f}_2|\ldots|\boldsymbol{f}_j|\ldots|\boldsymbol{f}_p]$ defines an identity matrix, $\boldsymbol{I}_{p\times p}$ and leads to the following proposition:

Proposition 1: The j^{th} product inventory $\boldsymbol{f}_j^T\boldsymbol{r}$ is an estimable function if and only if the j^{th} column of $\boldsymbol{P}_A$ is exactly equal to $\boldsymbol{f}_j$, i.e., $\boldsymbol{P}_A\boldsymbol{f}_j=\boldsymbol{f}_j$. The life cycle inventories for such functional units, and any arbitrary linear combination of such units, are estimable functions, and there is no need to do allocation to find their inventories.

Equivalently, one may check if $(\boldsymbol{I}-\boldsymbol{P}_A)\boldsymbol{f}=0$. Looking at the non-zero columns of $\boldsymbol{I}-\boldsymbol{P}_A$, one can learn about the additional data that might be useful so as to avoid allocation. When additional data cannot be acquired, these columns indicate the product inventories that are most susceptible to the influence of a chosen allocation criterion.

Many software packages such as MATLAB and R contain ready-made functions for computing a projection matrix as well as the pseudoinverse, $\boldsymbol{A}^+$. When the desired functional unit is a non-estimable parametric function, changing the question or the description of the production network as suggested by Heijungs and Frischknecht ((1998), p.331) leaves the original question unanswered. Rather than reliance on changing the original question that needed an answer to a question that can be answered with a given technology matrix, the LCA community may want to focus on identification and acquisition of additional information, so that the desired question can be satisfactorily addressed.

2.2.2 An allocation-based solution is choice of a specific generalized inverse

Any partitioning procedure must use some additional information or principle to overcome the rank deficiency of $\boldsymbol{A}$. If a functional unit of interest is not estimable in the original data, combining the data with some additional knowledge or stated rules, the original problem is replaced by a new problem in which every functional unit is estimable. We now show that the "unique" solution $\boldsymbol{r}_*$ of Eq. 7 is one which solves Eq. 5 (equivalent to Eq. 15) plus an additional set of *p-n* linear equality constraints called *side conditions*. In short, solutions obtained by partitioning are equality-constrained least squares solutions of Eq. 5. These side conditions could be written in the form

$$\boldsymbol{K}\boldsymbol{r}_* = \boldsymbol{m} \tag{21}$$

and they say something about the actual values or relative magnitudes of all non-estimable life cycle intensities. In general, the side conditions in Eq. 21 could be on any *p-n* non-estimable parametric functions alone so that the matrix $\begin{bmatrix}\boldsymbol{A}^T\\ \boldsymbol{K}\end{bmatrix}$ is of full rank. By virtue of partitioning the original data into virtual processes, both $\boldsymbol{K}$ and $\boldsymbol{m}$ are functions of allocation weights, and the side conditions are ultimately set by the choice of weights *w*. To see what these additional assumptions actually entail, we suggest applying the full-rank transformation defined below to Eq. 7 to recover the side conditions imposed by partitioning.

An important point to note is that Eq. 5 can be recovered from Eq. 7 by aggregating over virtual processes from the same parent process. Let the matrix $\boldsymbol{J}$ denote an "aggregator" matrix which sums over collections of virtual processes (appropriate rows of Eq. 7) derived from the same parent process. Let $\boldsymbol{D}$ denote a "differentiator" which takes differences within collections of virtual processes derived from the same parent process (differences of appropriate rows of Eq. 7). By taking sums and differences, the rows of $\boldsymbol{D}$ will be orthogonal to the rows of $\boldsymbol{J}$. Writing these operations into a single matrix we have

$$\boldsymbol{Q} = \begin{bmatrix}\boldsymbol{J}\\ \boldsymbol{D}\end{bmatrix}. \tag{22}$$

The upper partition of $\boldsymbol{Q}$ corresponds to the aggregation operation; it will recover the original, *n* equations in *p* unknowns. The lower partition of $\boldsymbol{Q}$ is used to obtain side conditions implied by partitioning. Applying the transformation, we have

$$\boldsymbol{Q}\boldsymbol{A}_*^T\boldsymbol{r} = \boldsymbol{Q}\boldsymbol{b}_* \tag{23}$$

which yields

$$\begin{bmatrix}\boldsymbol{A}^T\\ \boldsymbol{K}\end{bmatrix}\boldsymbol{r} = \begin{bmatrix}\boldsymbol{b}\\ \boldsymbol{m}\end{bmatrix}. \tag{24}$$

For the simple example in Section 2.1.3, these operations are demonstrated in Section 3.

Since the matrix $\boldsymbol{Q}$ is of full rank, a solution to Eq. 7 is also a solution of Eq. 24 the lower partition of which contains a set of side conditions on non-estimable product inventories. Under these constraints, a solution to Eq. 5 becomes unique. These constraints may be set quite arbitrarily and still produce a solution to the unallocated equations in the upper partition without any effect on the estimate of any of the estimable functions. Since both $\boldsymbol{K}$ and $\boldsymbol{m}$ in Eq. 24 are functions of the allocation weights assigned to virtual processes, *the side conditions are, in fact, set quite arbitrarily* as each different set of allocation weights generates a different set of side conditions.

Equivalently, a solution of Eq. 24 is also the simultaneous solution of the normal equations and side conditions:

$$\begin{bmatrix}\boldsymbol{A}\boldsymbol{A}^T\\ \boldsymbol{K}\end{bmatrix}\boldsymbol{r} = \begin{bmatrix}\boldsymbol{A}\boldsymbol{b}\\ \boldsymbol{m}\end{bmatrix}. \tag{25}$$

The unique solution $\boldsymbol{r}_*$ of *equivalent systems* of Eqs. 7, 24, and 25 satisfies the additional constraints *and* corresponds to a solution of the normal equations with a particular g-inverse denoted $(\boldsymbol{AA}^T)^-$. Hence, the solution is $\boldsymbol{r}_*$ as shown in Eq. 17. The vector $\boldsymbol{r}_*$, a solution derived by partitioning, is a solution of Eq. 5, and since $\boldsymbol{A}^T(\boldsymbol{AA}^T)^-\boldsymbol{AA}^T=\boldsymbol{A}^T$, we note that $(\boldsymbol{AA}^T)^-\boldsymbol{A}$ is a generalized inverse of $\boldsymbol{A}^T$.

Differing allocation principles may entail different weights, which in turn determine a different set of side conditions. This changes the values of all non-estimable product inventories, sometimes quite radically. For each vector of emissions or resources consumed and a choice of allocation weights, an allocation procedure is equivalent to a choice of some g-inverse, $\boldsymbol{G}$, of the rectangular technology matrix, $\boldsymbol{A}^T$.

Note that while the matrix $(\boldsymbol{AA}^T)^-(\boldsymbol{AA}^T)$ in Eq. 26 below determines a solution to the normal equations, the expression *is not* an identity matrix so that

$$\begin{aligned}\boldsymbol{r}_* &= \left(\boldsymbol{AA}^T\right)^-\boldsymbol{Ab} \\ &= \left(\boldsymbol{AA}^T\right)^-\boldsymbol{AA}^T\boldsymbol{r}\neq\boldsymbol{r}.\end{aligned} \tag{26}$$

This means that the solution chosen by allocation, $\boldsymbol{r}_*$, is a *biased estimate* of $\boldsymbol{r}$, the fixed but unknown list of life cycle intensities for each product with respect to some emission, $\boldsymbol{b}$. Unless a particular inventory is already estimable from the given data, different allocation procedures will induce different biases into the non-estimable components of $\boldsymbol{r}_*$, which may not be remotely close to one another. *This is the source of discrepancy between any two partitioning criteria*, and it has significant implications for current practices in LCA. Estimability of the life cycle inventories for every product in the technological system can be achieved only through the acquisition of more information or data. Principles from the considerable literature on statistical design of experiments (Dean and Voss 1999) may offer some insights to guide this acquisition, so that at least the product inventory of interest is estimable.

We wish to emphasize that any solution derived by partitioning is a non-unique solution of Eq. 5 corresponding to some g-inverse of the technology matrix. Each partitioning allocation corresponds to one particular solution in the set of all possible solutions $R_{\boldsymbol{b}}$ that was shown in Eq. 18. When weights are constrained in the interval [0,1] and sum to 1 for virtual processes from the same parent process, this set of all possible partitionings of the data constitute a subset $C_{\boldsymbol{b}}$ of all possible solutions.

2.2.3 Resolution of Heijungs and Suh's conjecture 1

The resolution of a long-standing conjecture follows directly from the discussion of the previous section. Related to the previous work in the study of Heijungs and Frischknecht (1998), Heijungs and Suh conjectured about the relationship between inventories derived by partitioning and by solving for scale factors in the overdetermined system Eq. 1 , using the Moore-Penrose inverse:

"Conjecture 1: A system with a technology matrix which has more rows [products] than columns [processes] can be solved exactly in either of two ways: with the pseudoinverse of the technology matrix

$$\boldsymbol{s} = \boldsymbol{A}^+\boldsymbol{f} \tag{27}$$

or with

$$\boldsymbol{s}_* = \boldsymbol{A}_*^{-1}\boldsymbol{f}_* \tag{28}$$

where $\boldsymbol{A}_*$ is a square and invertible matrix that is derived from the technology matrix $\boldsymbol{A}$ and $\boldsymbol{f}_*$ is derived from $\boldsymbol{f}$ by means of eliminating rows of $\boldsymbol{A}$ and $\boldsymbol{f}$ that correspond to cut-off, by adding columns to $\boldsymbol{A}$ to account for so-called avoided processes, and by replacing columns of $\boldsymbol{A}$ that represent multifunctional processes by a number of monofunctional processes." ((Heijungs and Suh 2002), p. 91)

Several ideas are tied up in this statement, and all of them require estimability of the desired functional unit. First, any overdetermined linear system can be solved exactly only if it is a consistent system of equations. By the very definition of estimability (see definition 2), an exact solution of the system Eq. 1 is required in order for the inventory for the stated functional unit, $\boldsymbol{f}^T\boldsymbol{r}$, to be an estimable function. When the system Eq. 1, overdetermined in products, is still consistent, those n scale factors which solve the system exactly exist and the Moore-Penrose inverse could be used to solve for them, *as could any other generalized inverse of* $\boldsymbol{A}$*, some of them corresponding to partitioning allocations*. Proposition 1 of Section 1 can be used to identify a basis set for all possible estimable functions from the unallocated data.

Another interesting phenomenon has been observed in terms of the scale factors in allocated and unallocated systems that sometimes the scale factors of virtual processes take on the same value as their parent process in the unallocated system (cite Heijungs and Suh 2002, p. 71). We show in Section 3 of the supporting information that this is also a consequence of estimability and that the phenomenon occurs when the scale factors for estimable functional units of the unallocated technology matrix are propagated through the aggregator matrix.

Based on the findings of Sections 2.2.1 and 2.2.2 and the discussion above, the conjecture should now be replaced by the following statements.

I. There exists a partitioning of the data "into monofunctional processes" which corresponds to some choice of generalized inverse of the technology matrix.
II. When the desired functional unit is estimable, its inventory is invariant to choice of partitioning criterion.

While Heijungs and Suh specifically conjectured about exact solutions, statement *I* is more generic in that it even links approximate solutions of Eq. 5 with the outcome by some partitioning. We note that Eq. 27 above represents ordinary least squares (OLS) allocation pursued by Marvuglia et al. (2010). *The least squares allocations pursued therein are equivalent to partitioning allocation.* In the partitioned matrices corresponding to OLS and the other least squares allocations (data least squares (DLS) and total least squares (TLS)), the weights assigned to virtual processes from the same parent process will sum to 1, but they are not guaranteed to lie in the interval [0,1]. In such cases, it is as if *the inputs of the parent process become outputs in some virtual processes while being produced in excess in others*. This makes the reporting and interpretation of such allocations problematic as we show in the next section.

3 The illustrative example continued

3.1 Distinguishing estimable and non-estimable functional units

For the matrix

$\boldsymbol{A}$ in Eq. 9, $rank(\boldsymbol{A})=3$, there are three linearly independent, estimable functional units; inventories for arbitrary linear combinations of those functional units are also estimable. The projection matrix for the given technology matrix is computed by using Eq. 19, and

$$\boldsymbol{P}_A = \begin{bmatrix} .5 & .5 & 0 & 0 \\ .5 & .5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{29}$$

An inspection of this projection matrix reveals that there would be no need for allocation if the desired functional units for study were 1 L of fuel oil or 1 kWh of electricity since the projections for those functional units (the third and fourth columns of P_A) are exactly equal to $\boldsymbol{f}_3$ and $\boldsymbol{f}_4$, respectively. Furthermore, any multiple or linear combination of those two product inventories would be estimable and would be invariant to the choice of partitioning criterion. The first two columns of $\boldsymbol{P}_A$, however, indicate that inventories for a single kilogram of waste 1 or for 1 kg of waste 2 are not estimable as their projections are not identical to $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$, respectively. More than that, the two co-products are indistinguishable in the given data. What can be estimated from the given data is *an inventory for a fixed mix of the two wastes*: one unit of waste 1 for every unit of waste 2. Consequently, the third linearly independent estimable basis function is $[1\ 1\ 0\ 0]\boldsymbol{r}= r_1+r_2$, the sum of inventories for wastes 1 and 2, which are both produced by the separation unit. This also confirms the observations made about Eq. 13.

3.2 The side condition and its interpretation

Note that each intensity vector $\boldsymbol{r}_*$ is a particular solution of Eq. 5, which also satisfies one additional side condition determined by the choice of allocation weight. We now show how the side conditions for an arbitrary partitioning allocation can be derived. An arbitrary value assigned to some non-estimable function, say the difference

$$[1\ -1\ 0\ 0]\boldsymbol{r} = r_1-r_2 \tag{30}$$

would provide a unique answer to the inventory involving these products, but, given infinitely many choices for this difference, interpretation of this allocation based inventory must be done *with respect to the side condition*.

The side conditions corresponding to an allocation weight can be derived by examining Eq. 23; for this example, the transformation matrix takes the form

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{J} \\ \boldsymbol{D} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & 0 \end{bmatrix} \tag{31}$$

where $\boldsymbol{J}$ aggregates over the first two rows of Eq. 7 to recover the original three equations in the four unknown product intensities, and the lower partition $\boldsymbol{D}$ takes differences of the equations representing virtual processes. Defining the product $\boldsymbol{D}\boldsymbol{A}_*^T \equiv \boldsymbol{K}$ and $\boldsymbol{D}\boldsymbol{b}_* \equiv \boldsymbol{m}$ indicates the side condition as a function of the allocated structures, i.e., the lower partition of Eq. 24.

For any vector of interventions $\boldsymbol{b}$, this results in

$$\boldsymbol{K} = [\,1 \quad -1 \quad 0 \quad -2(2w-1)\,] \tag{32}$$

and

$$\boldsymbol{m} = b_1(2w-1) \tag{33}$$

where b_1 indicates the emissions measured at first process (from which both types of waste originate) in the unallocated system. Using Eq. 21, the side condition simplifies to

$$r_1-r_2 = (2r_4 + b_1)(2w-1). \tag{34}$$

Substituting the choice of w in Eq. 13 produces the inventories by traditional means. Substitution of w in Eq. 34 reveals the side condition associated with either allocation. For this example, allocation based on the mass of the separated waste stream would yield a 50–50 split of inputs, emissions, wastes, and resources consumed. When the 50–50 split is assumed in this example (i.e., when $w=0.5$), the side condition reduces to the assumption that, i.e., the life cycle effects of waste 1 and

waste 2 are equal. Since b_1 was arbitrary in Eq. 34, an equal allocation between the two co-products, the assumption implicitly made is that the life cycle effects of each co-product are identical. In statistics jargon, each unit of waste 1 serves as an *alias* for each unit of waste 2 and it is assumed that their effects are indistinguishable.

When $w \neq 1-w$, the side condition in Eq. 34 also depends on interventions due to the separation process (b_1) and the life cycle intensity of inputs of electricity (r_4) with respect to b_1. If the prices per unit of wastes 1 and 2, which might be recovered for some valuable use outside the system, were $4 and $1, respectively, then economic allocation would yield an 80–20 split of inputs and interventions. Interpreting, Eq. 34, what we report as "economic allocation" is really an assumption about the difference of the life cycle effects of the two types of waste, where the magnitude of that difference (some 50.4 g for CO_2, 47.4 g for NO_x) has been *determined by measured emissions at the first process, the inventory of another input within the network boundary and the weight assigned to each waste stream.* Similarly, a 100 % assignment of emissions to waste 1, assumes these differences are 84 g CO_2 and 79 g NO_x, respectively.

Figure 2 depicts a variety of the possible outcomes for the life cycle inventory of CO_2 due to waste 1 as a function of the allocation weight assigned to waste 1. The dashed line represents inventories obtained from all possible solutions ($R_{\boldsymbol{b}}$), without restriction on the partitioning weight, literally

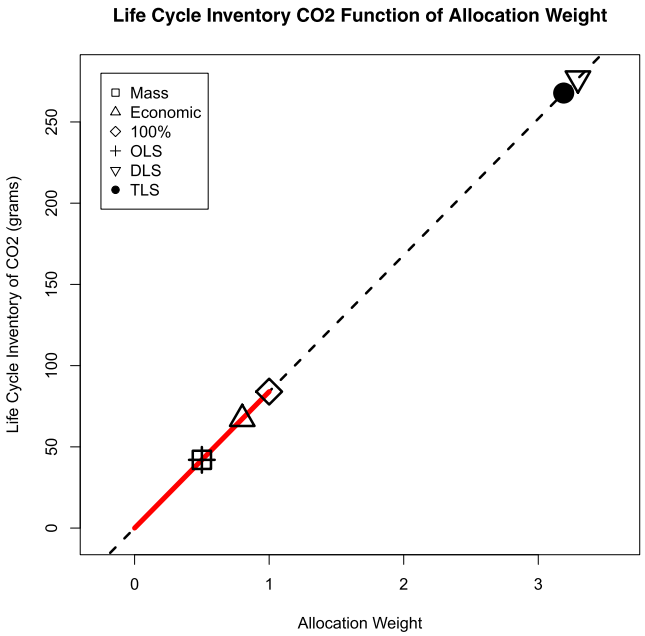

Fig. 2 Inventory of CO_2 as a function of (equivalent) partitioning weight. The *dashed line* represents a set of all possible solutions. The *solid red line* represents subset of solutions obtained by partitioning with $0 \leq \boldsymbol{w} \leq 1$. In this case, OLS and mass-based allocation provide equivalent results. Note that equivalent partitioning weights for DLS and TLS are greater than 1

any real number is a possible inventory outcome. A plot for NO_x, or for any type of intervention one might be able to measure, would look quite similar. The solid red line depicts the set $C_{\boldsymbol{b}}$ of inventories of CO_2 obtained by standard partitionings which restrict $0 \leq w \leq 1$. The set of standard partitioning outcomes is clearly subset of all possible solutions. In order to illustrate where they might fall relative to one another on this continuum of possible solutions, we have plotted six separate allocations as points, including equivalent partitionings for the three least squares methods pursued in the study of Marvuglia et al. (2010). More analysis on the case study data published in the study of Marvuglia et al. (2010, 2012) is in Cruze et al. (2014).

Figure 2 illustrates that looking for agreement among these few chosen outcomes is not necessarily meaningful. While mass and OLS allocations coincide, the equivalent partitionings for DLS and TLS assign more than 300 % share of parent process inputs and emissions to waste 1, and a −200 % share to waste 2 and produce an inventory more than six times the magnitude of mass-based allocation. Schematics representing the mass-based virtual network and the equivalent TLS virtual network are given in the supporting information. Note that in the TLS network, the first virtual process takes in more electricity and emits more CO_2 per unit of operation time than the parent process, and the second virtual process produces excess electricity and acts as a sink for CO_2. Taking either of the DLS or TLS inventories at face value requires us to believe that waste 2 is somehow capable of producing electricity from CO_2, and of avoiding 183 to 192 g of CO_2 emissions, despite the fact that *data* only tells us that the combined life cycle emissions for the two wastes, $r_1+r_2=84\,\mathrm{g}$ of CO_2. The use of least squares techniques is not a satisfactory solution to the allocation problem, and reporting and acting on any of these outcomes requires us to *believe that the associated side condition is at least plausible*.

Without ever resorting to a partitioning allocation, one could specify a side condition in the form of an assumption about a single free parameter, in this instance, either of the two waste products. This relates to the "duality" described in the study of Cherubini et al. (2011), which asserts that system boundary expansion could be represented as a partitioning allocation. In their view, the act of system boundary expansion may not have entailed the literal acquisition of more measurements on other processes but additional information about some other products whose uses may have been offset by one of the co-products.

For the sake of illustration, assume that the life cycle effects due to 1 kg of waste 1 are of interest, but it is known that each kg of waste 2 can offset the use of 2 kg of some other product outside the system boundary. Previous studies suggest that life

cycle releases for this other product total 14 g CO_2 /kg product and 10 g NO_x/kg product. By setting

$$r_2^{CO_2} = \frac{2\ \text{kg outside product}}{1\ \text{kg waste 2}} \times \frac{14\ \text{g } CO_2}{1\ \text{kg outside product}} = \frac{28\ \text{g } CO_2}{\text{kg waste 2}} \quad (35)$$

$$r_2^{NO_x} = \frac{2\ \text{kg outside product}}{1\ \text{kg waste 2}} \times \frac{10\ \text{g } CO_2}{1\ \text{kg outside product}} = \frac{20\ \text{g } CO_2}{\text{kg waste 2}} \quad (36)$$

we assert that

$$r_1^{CO_2} = 84-28 = 56\ \text{g } CO_2 \quad (37)$$

$$r_1^{NO_x} = 79-20 = 50\ \text{g } NO_x. \quad (38)$$

The inventories presented in Eqs. 37 and 38 are made under the side conditions (an assumption about the life cycle intensity of waste 2) shown in Eqs. 35 and 36, respectively. This is often the goal of system boundary expansion or substitution, to derive a credit or amount of offset that a co-product provides when it displaces products elsewhere. If this sort of information is available, then it can be incorporated easily through the use of side conditions without having to acquire additional process units to make a square invertible matrix, an apparent limiting factor in the application of system boundary expansion.

4 Conclusions

By convention, life cycle assessment is usually conducted in the column space of the technology matrix. That is, the problem is usually thought of as selection of scaling factors consistent with a desired functional unit and interpreting the inner product of environmental interventions with these scaling factors as the inventory of interest. Results found in this manner, however, cannot be divorced from the underdetermined system of equations represented by Eq. 5. Life cycle inventory could just as easily be thought of as *a case of regression through the origin*, and through this lense, the main issue of allocation by partitioning in process based life cycle inventories becomes clear: the inventory quantities we wish to interpret are often non-estimable parametric functions. This means that they are (1) not unique and (2) greatly affected by any additional information brought to bear on extracting a solution.

The theory of less-than-full-rank linear models provides a *general explanation* as to why two inventories for the same functional unit sometimes differ so greatly. These concepts also encompass the least squares techniques presented in the study of Marvuglia et al. (2010, 2012). Approaching life cycle inventory from the context of an underdetermined system of equations, it is clear that only certain functions of inventories for individual products will be estimable functions and all others will be determined by the side conditions imposed by the allocation.

While partitioning may offer a prescribed protocol for obtaining "the inventory," it does not guarantee that the additional information is incorporated in a truly meaningful manner. For instance, one could make an appeal to conservation principles and choose either mass or energy as a partitioning criterion, but if the weights differ between the two criteria, the resulting inventories for non-estimable functional units may differ, perhaps even greatly. On one hand, we expect that both mass and energy balances must hold within the given system boundary. On the other hand, we may arrive at two very different understandings of embodied energy consumption and life cycle emissions based on one allocation principle or the other. Since each method of partitioning entails a certain set of side conditions, it is difficult to imagine that any one criterion (mass, energy, economic valuation,…) will yield meaningful inventories across all life cycle studies. The challenge is to determine whether the relationships among product inventories described by the associated side conditions are consistent or at least plausible given the additional information we wish to bring to bear on the problem.

Rather than arbitrarily setting side conditions through a choice of allocation weight, one could specify the required number of side conditions on non-estimable parameters directly. These might reflect expert knowledge or values from previous studies about the intensities of a few of the products in the system. For example, a condition might reflect the belief that the life cycle intensity of one product is several times greater than another within the same system, or it might be a fixed value assigned to the intensity of some input, say electricity, based on previously published or industry-standard values.

Since the inventory phase represents the second step in an LCA study, the subsequent impact assessment and interpretation phases are driven by any and all assumptions made in the inventory step. Consequently, allocation is, as Marvuglia, Cellura and Heijungs put it, "the Achilles heel" of life cycle assessment ((Marvuglia et al. 2010), p.1036). Since the number of solutions to choose from is infinite, even consensus among a few chosen allocation methods would not necessarily lend validity to an LCA study. Squareness of the technology matrix is irrelevant; when the inventory for the product of interest is estimable, it can be found in the unallocated system. Instead, rank deficiency of the technology matrix and non-

identifiability of the product inventories of interest are the key issues. Rather than relying on partitioning, a better approach would be to seek out additional data or expert judgment so that at least the parameter (single product inventory) of interest becomes estimable. To the extent that system boundary expansion may be synonymous with the acquisition of more observations (processes), this approach should be preferred to partitioning, regardless of the choice of partitioning criterion (Weidema 2001; Weidema and Schmidt 2010). Furthermore, replicates showing variation in combination of inputs need not be avoided. When the additional data simply cannot be acquired, solutions by carefully and explicitly stated side conditions may be more meaningful and easier to disclose than the solutions derived by partitioning allocation.

References

Ben-Israel A and Grenville T (2003) Generalized inverses: theory and applications. Springer-Verlag

Cederberg C, Stadig M (2003) System expansion and allocation in life cycle assessment of milk and beef production. Int J Life Cycle Assess 8:350–356

Cherubini F, Strømman AH (2011) Life cycle assessment of bioenergy systems: state of the art and future challenges. Bioresour Technol 102:437–451

Cherubini F, Strømman AH, Ulgiati S (2011) Influence of allocation methods on the environmental performance of biorefinery products—a case study. Resour Conserv Recy 55:1070–1077

Cruze NB, Goel PK, Bakshi BR (2014) Revisiting least squares techniques for the purposes of allocation in life cycle inventory. Int J Life cycle Assess accepted. doi:10.1007/s11367-014-0771-9

Curran MA (2007) Studying the effect on system preference by varying coproduct allocation in creating life-cycle inventory. Environ Sci Technol 41:7145–7151

Dean A and Voss D (1999). Design and analysis of experiments. Springer

DeGroat R, Dowling E (1993) The data least squares problem and channel equalization. IEEE Trans Signal Process 41(1):407–411

Doluweera G, Jordaan S, Moore M, Keith D, Bergerson J (2011) Evaluating the role of cogeneration for carbon management in Alberta. Energy Policy 39:7963–7974

Flysjö A, Cederberg C, Henriksson M, Ledgard S (2011a) How does co-product handling affect the carbon footprint of milk? Case study of milk production in New Zealand and Sweden. Int J Life Cycle Assess 16:420–430

Flysjö A, Henriksson M, Cederberg C, Ledgard S, Englund J-E (2011b) The impact of various parameters on the carbon footprint of milk production in New Zealand and Sweden. Agr Syst 104(6):459–469

Golub G, Kahan W (1965) Calculating the singular values and pseudoinverse of a matrix. J Soc Ind Appl Math B Met 2:205–224

Harville DA (1997) Matrix algebra from a statistician's perspective. Springer

Heijungs R, Frischknecht R (1998) A special view on the nature of the allocation problem. Int J Life Cycle Assess 3(5):321–332

Heijungs R, Suh S (2002) The computational structure of life cycle assessment. Kluwer Academic Press, Dordrecht

Kalaba R, Xu R, Feng W (1995) Solving shortest length least-squares problems via dynamic programming. J Optimiz Theory App 85:613–632

Kutner MH, Nachtsheim CJ, and Neter J (2004). Applied Linear Regression Models (Fourth ed.). McGraw-Hill Irwin

Lay DC (2003) Linear algebra and its applications. Pearson Education

Marvuglia A, Cellura M, Heijungs R (2010) Toward a solution of allocation in life cycle inventories: the use of least-squares techniques. Int J Life Cycle Assess 15:1020–1040

Marvuglia A, Cellura M, Pucci M (2012) A generalization of the orthogonal regression technique for life cycle inventory. Int J Agr Environ Inf Syst 3:51–71

Meyer CD (2001). *Matrix analysis and applied linear algebra*. Society for Industrial and Applied Mathematics

Rao CR (1962) A note on a generalized inverse of a matrix with applications to problems in mathematical statistics. J Roy Stat Soc B Met 24(1):152–158

Ravishanker N and Dey DK (2002) A first course in linear model theory. Chapman and Hall/CRC

Searle SR (1965) Additional results concerning estimable functions and generalized inverse matrices. J Roy Stat Soc B Met 27(3):486–490

Silalertruksa T, Gheewala SH (2011) Long-term bioethanol system and its implications on GHG emissions: a case study of Thailand. Environ Sci Technol 45:4920–4928

Van Huffel S and Vandewalle J (Eds.) (1991) The total least squares problem: computational aspects and analysis. Soc Ind and Appl Math

Wang M, Huo H, Arora S (2011) Methods of dealing with co-products of biofuels in life-cycle analysis and consequent results within the U.S. context. Energy Policy 39:5726–5736

Wardenaar T, van Ruijven T, Mendoza Beltran A, Vad K, Guinée J, Heijungs R (2012) Differences between LCA for analysis and LCA for policy: a case study on the consequences of allocation choices in bio-energy policies. Int J Life Cycle Assess 17:1059–1067

Weidema B (2001) Avoiding co-product allocation in life-cycle assessment. J Ind Ecol 4(3):11–33

Weidema BP, Schmidt JH (2010) Avoiding allocation in life cycle assessment revisited. J Ind Ecol 14:192–195